

Results Submission 3

Caroline Sullivan, Joe Thompson, Sasha Porter, Genevieve Purcell

2024 is a historical year with significant implications of the presidential election. Given that it is the season for a pivotal presidential election, we wanted to use machine learning tactics to predict the election outcome for the state of Virginia. Our prediction question is as follows:

Considering voting data from the 2000, 2004, 2008, 2012, 2016, and 2020 presidential elections, how significant are historical voting trends in accurately predicting presidential election outcomes for specific counties in the state of Virginia?

We found that, by using historical voting trends such as the trend of Democratic and Republican vote share, our model predicted the results of the 2024 Presidential Election for each county in Virginia with 78% accuracy.

In this discussion, we will highlight historical voting trends for the state of Virginia in parts I and II. Then we will discuss the results we derived from our logistic regression model in part III. To conclude, in part IV we will compare our findings and predictions to the actual results of the 2024 presidential election for the state of Virginia.

I. Historical Voting Trends

Figure 1:

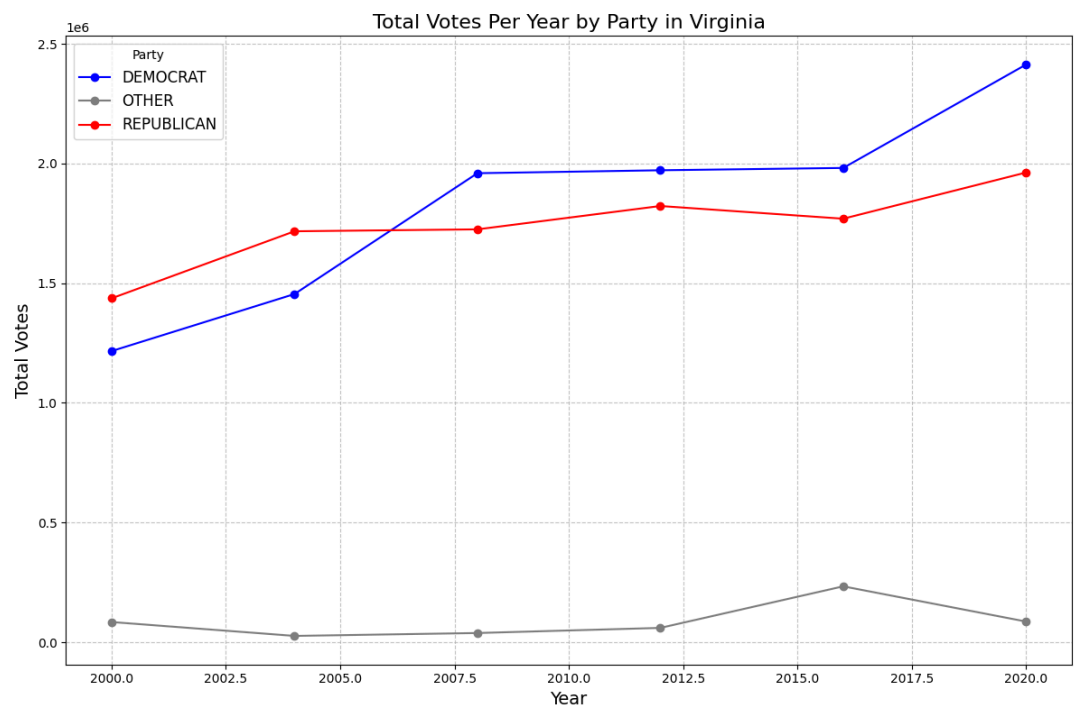


Figure 1 illustrates voter turnout trends over the past 20 years. The blue line represents the number of votes received by the Democratic party each year, while the red line represents the number of votes received by the Republican party each year. The grey line aggregates votes for third-party candidates not affiliated with a major party.

The data reveals a significant upward trend in Democratic voter turnout with a notable spike in Democratic votes in 2020, shown as the blue line. This is likely due to higher Democratic engagement and possibly demographic shifts favoring the party. In contrast, the red line is showing the Republican votes and how they had moderate growth from 2000 to 2004, but since then have stagnated with only slight fluctuations. The grey line shows that third-party votes have had minimal impact overall, with a slight increase in 2016 which likely represents a third-party candidate gaining traction in a specific election.

Overall, the total number of votes has increased steadily, reflecting growing voter participation in recent elections. This trend may be driven by heightened national political polarization, which tends to boost voter engagement. Contributing factors include population growth and demographic changes, particularly in Virginia. Over the past few decades, the state has seen significant population increases and greater diversity, especially in urban and suburban areas, which traditionally lean Democratic.

II. Past Results

Figure 2:

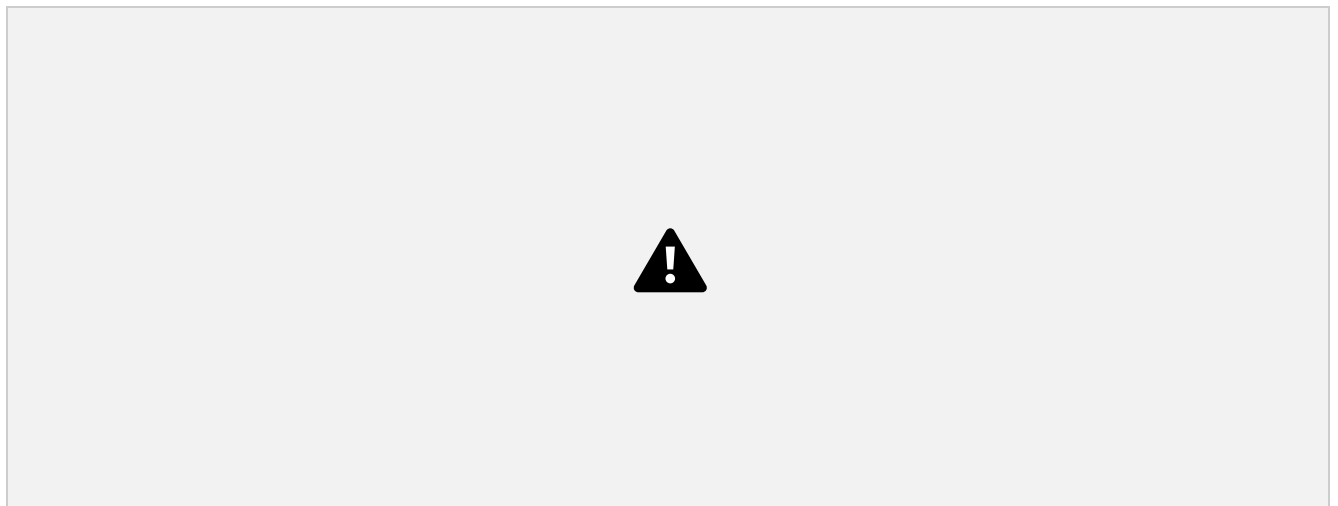


Figure 2 is a video of an [interactive map](#) of Virginia showing the difference in county voting results over the 20 year period. Counties are colored by who won their majority vote, with red being Republican and blue being Democratic.

There are many counties who have stayed the same color (or same majority vote) over the past 20 years, but there are many who have changed their color. In 2000, there were a lot

more red counties than in 2020, and this is likely due to the change in Virginia's population and move to more urban and suburban areas. Election-specific dynamics is also a big factor in what the majority vote is in each county as many people may be partial to a specific candidate.

III. Results

Based on our [logistic regression model](#), we ran a regression analysis to find the following coefficient results:

Figure 3: Coefficient Results

	variable	coefficient
0	democratic_share_trend	-0.357703
1	republican_share_trend	-3.409907
2	other_share_trend	-0.290734
3	turnout_change	0.241128
4	normalized_margin_of_victory	-0.178615

To predict the election outcomes in Virginia counties, we derived feature variables based on the historical voting trends found in Virginia's voting data. These variables consisted of vote share trends by party (Democratic, Republican, and other), the trend of voter turnout, and the margin of victory. With the use of these feature variables, our model resulted in a 77% training accuracy and 75% testing accuracy. When running the regression on testing data, the model was found to have a 77% precision when classifying a record as 0 and a 62% precision when classifying it as 1. This indicates that it was more accurate in predicting a Republican win over a Democratic win. Furthermore, there were 32 false negatives (Democratic wins misclassified as Republican) compared to 9 false positives (Republican wins misclassified as Democratic), suggesting the model more frequently underestimated Democratic victories.

The model's higher accuracy in predicting Republican outcomes can be explained in the coefficients for our feature variables. The results suggest that the Republican vote share trend impacts the accuracy of our model the most, with a coefficient of -3.4. Thus, a significant upward trend in the percentage of votes a Republican candidate receives is a strong predictor of the likelihood of a Democratic loss. We can also see that the Democratic vote share trend has a coefficient of -0.36, implying that a rising trend in Democratic vote share is less likely to have a Democratic candidate win. This might seem counterintuitive, but it is important to note that an increasing Democratic trend does not directly imply a Democratic win in that county. Because these variables have large negative coefficients, our model leans towards a lower likelihood of a Democratic win.

IV. Comparison to Actual Results

After comparing the results of our model to the actual results of the 2024 presidential election, the model had a 78% accuracy when predicting the results of the voting data of the 2024 presidential election based on trends from prior elections. One reason for this discrepancy

is that voters' tendencies can change as a result of external variables, suggesting that they can flip between years. Some issues can appear more prevalent and be more likely to influence voting decisions between election cycles that our model can not account for. The model's performance highlights its potential for identifying voting patterns, though there remains room for improvement in refining its predictive capabilities by incorporating additional factors such as demographic shifts, voter turnout variability, and emerging political dynamics.

Figure 4: 2024 Presidential Election Predicted Results Map

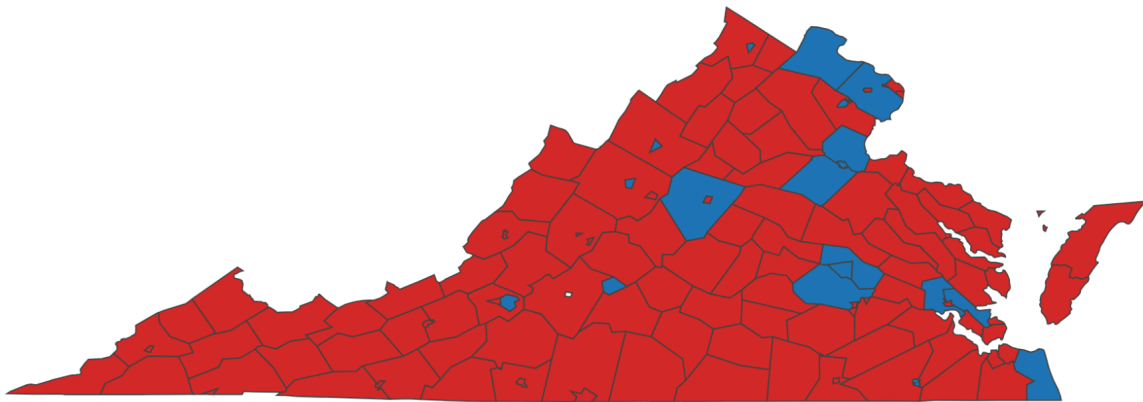


Figure 5: 2024 Presidential Election Actual Results Map

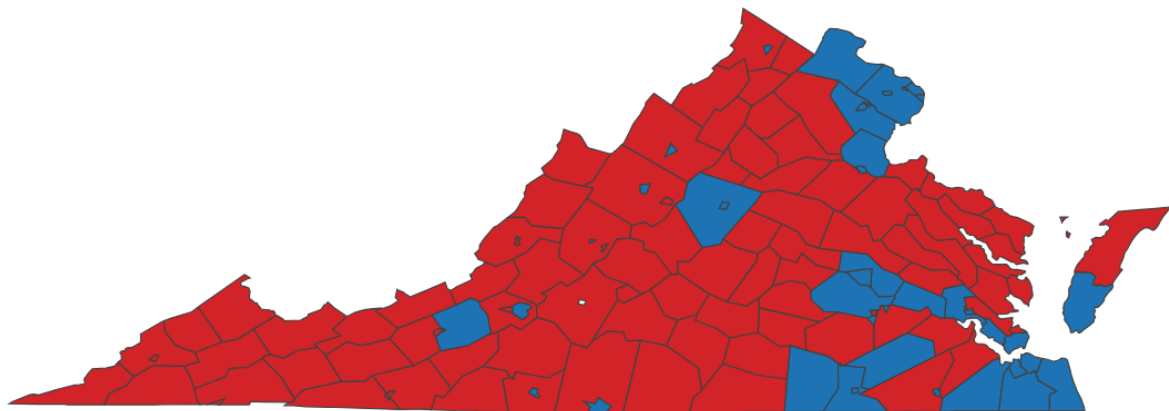


Figure 4 shows the 2024 presidential election results that our logistic regression model predicted. The map in Figure 4 shows a strong prediction of Republican dominance (red) across

most counties, with only a few urban or suburban counties predicted to vote Democratic (blue). Figure 5 shows the actual 2024 presidential election results. In the actual results map some counties predicted to vote Republican appear blue, indicating a shift in party support. There is also a noticeable increase in Democratic representation, particularly in regions where urban and suburban areas are located. Generally, we were correct in predicting the Greater Richmond area and Northern Virginia counties. We were incorrect in predicting the coastal region outside of Virginia Beach, such as the Eastern Shore and some areas along the James River. These locations voted Democratic rather than Republican as we predicted. The same is true for Brunswick, Greenville, and Sussex counties.

V. Results Conclusion

Our analysis demonstrates the utility of historical voting trends in predicting election outcomes, while acknowledging limitations of the model. Using a logistic regression model, we achieved a 78% accuracy rate in forecasting the results of the 2024 presidential election for Virginia counties by leveraging features such as party vote share trends, voter turnout changes, and margins of victory. Overall, the model did well in predicting Republican outcomes but struggled slightly with Democratic wins. Comparing our predictions to actual results revealed areas of strength, such as correctly forecasting outcomes in Northern Virginia and Greater Richmond, and areas for improvement, particularly in regions like the Eastern Shore and the James River area. These discrepancies underscore the importance of accounting for external variables, demographic shifts, and emerging political dynamics in future iterations of our model.